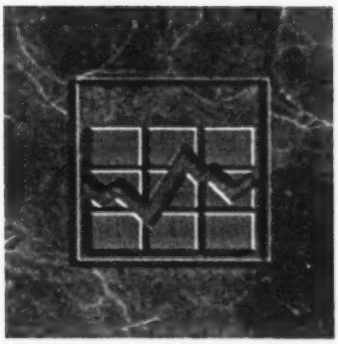
N° 62F0026M au catalogue — N° 001
ISSN 1708-8887
ISBN 978-1-100-97523-8

Document de recherche

Série de documents de recherche sur les dépenses des ménages

Enquête sur les dépenses des ménages de 2009 : indicateurs de qualité des données



Division des méthodes d'enquêtes auprès des ménages
Immeuble R.-H.-Coats, Ottawa, K1A 0T6

Téléphone: 613-951-7355

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca. Vous pouvez également communiquer avec nous par courriel à infostats@statcan.gc.ca ou par téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

Centre de contact national de Statistique Canada

Numéros sans frais (Canada et États-Unis) :

Service de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Télécopieur	1-877-287-4369

Appels locaux ou internationaux :

Service de renseignements	1-613-951-8116
Télécopieur	1-613-951-0581

Programme des services de dépôt

Service de renseignements	1-800-635-7943
Télécopieur	1-800-565-7757

Comment accéder à ce produit

Le produit nº 62F0026M au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de parcourir par « Ressource clé » > « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens ».

Statistique Canada
Division des méthodes d'enquêtes auprès des ménages

Enquête sur les dépenses des ménages de 2009 : indicateurs de qualité des données

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2011

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Juin 2011

N° 62F0026M au catalogue
ISSN 1708-8887
ISBN 978-1-100-97523-8

Périodicité : hors série

Ottawa

This publication is available in English upon request.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Ce document a été préparé par :

Georgina House
Denis Malo
Marie-Hélène Miville
Sylvain Nadon
Johanne Tremblay

de la Division des méthodes d'enquêtes auprès des ménages (DMEM)

Les auteurs remercient l'équipe responsable de l'Enquête sur les dépenses des ménages de la Division de la statistique du revenu (DSR) pour leur collaboration dans la préparation de ce document.

Table des matières

Faits saillants **6**

Introduction **9**

1. Les erreurs d'échantillonnage **10**

1.1 Les mesures de l'erreur d'échantillonnage 10
1.2 Les coefficients de variation 11
1.3 Modèle pour dériver une approximation du CV 12
1.4 Suppression des données non fiables dans les tableaux d'estimations 14

2. La non-réponse **14**

2.1 Le taux de réponse, les taux de non-réponse et les taux de vacance 15
2.2 La non-réponse selon le niveau d'urbanisation 17
2.3 La non-réponse selon les strates de revenu 18
2.4 L'ajustement pour la non-réponse 18

3. Les erreurs de couverture **19**

3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement 20
3.2 L'ajustement au niveau de la population et des ménages 22

4. Les erreurs de réponse **23**

5. Les erreurs de traitement **24**

5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale, provinciale et territoriale 25

Références **31**

Annexe A Notation algébrique **32**

Faits saillants

Les erreurs d'échantillonnage

- Les coefficients de variation (CV) de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,4 % et 2,3 % selon les provinces. Le CV à l'échelle nationale est de 1,0 %. Les CV sont plus élevés dans les territoires, soit 5,1 % au Yukon, 7,6 % dans les Territoires du Nord-Ouest et 6,0 % au Nunavut.

- Les coefficients de variation (CV) de l'estimation de la moyenne pour les différentes catégories de dépenses sommaires sont pour la plupart inférieurs ou égaux à 2,4 % à l'échelle nationale et généralement inférieurs à 5,0 % à l'échelle provinciale. Les résultats concernant les caractéristiques des logements et de l'équipement ménager sont similaires. La taille de l'échantillon étant moindre à l'Île-du-Prince-Édouard et dans les territoires, les CV ont tendance à être plus élevés que ceux des autres provinces.

La non-réponse

- Le taux de réponse final est de 64,5 %. Les taux de réponse provinciaux et territoriaux s'échelonnent de 59,6 % en Ontario à 77,9 % pour le Nunavut.

- Le taux de non-réponse est de 35,5 %. La non-réponse est due aux refus (21,7 %), aux ménages qu'on n'a pas réussi à contacter (10,8 %) et aux ménages dont les données ont été considérées inutilisables (3,0 %).

- Le taux de non-réponse a tendance à croître avec le niveau d'urbanisation. On observe un taux de non-réponse de 29,5 % en milieu rural, et un taux de 38,3 % dans les centres de population d'un million d'habitants et plus.

- L'analyse des taux de non-réponse dans les strates d'aires géographiques à revenu élevé provenant du plan d'échantillonnage révèle que le taux dans les strates à revenu élevé (40,9 %) est plus élevé que celui observé dans les autres strates (35,0 %). Le taux de refus des strates à revenu élevé (24,7 %) est supérieur à ce qui est observé pour les autres strates (21,5 %).

Les erreurs de couverture

- On observe un sous-dénombrement des ménages de 12,2 % à l'échelle nationale.

- Le sous-dénombrement des personnes se situe à 16,1 % à l'échelle nationale (excluant les terrritoires).

Les erreurs de traitement reliées à l'imputation

i) Des variables de dépenses

- Il a fallu imputer au moins une variable de dépenses[1] pour 58,9 % des ménages à l'échelle nationale. La section du questionnaire portant sur les dépenses reliées aux services de communication au foyer (téléphone, téléphone cellulaire et accès à Internet), aux services de télédistribution (câble), aux services de diffusion par satellite et aux services de sécurité contribue fortement à ce taux. Depuis 2004, on accepte qu'un ménage répondant fournisse seulement les dépenses totales d'un forfait (services groupés), tout en indiquant quels services sont inclus. Environ 45 % des ménages utilisables ont nécessité l'imputation d'au moins un de ces six services.

- Le taux global d'imputation en excluant ces six services est de 29,0 % à l'échelle nationale. À l'échelle provinciale et territoriale, les taux d'imputation varient entre un minimum de 14,4 % pour le Nunavut et un maximum de 34,7 % pour le Québec.

- Les valeurs imputées contribuent à 43,8 % de l'estimation du total des dépenses de services de télédistribution et à 52,8 % de l'estimation du total des dépenses de services d'accès à Internet. Les taux d'imputation parmi les ménages déclarant des dépenses à l'un ou l'autre de ces deux services sont respectivement de 44,5 % et 58,5 %. Ces taux plus élevés sont principalement dus au fait qu'une proportion importante des forfaits incluait ces deux services parmi les ménages qui ont déclaré avoir payé pour un forfait.

- Près de 10 % des individus ont nécessité l'imputation de variables sur les dépenses vestimentaires mais pour la très grande majorité, le total était fourni par le répondant et seules les composantes ont été imputées.

- Environ 16 % des personnes âgées de 15 ans et plus ont nécessité l'imputation d'au moins une des variables de la section sur les impôts personnels, sécurité et dons en argent.

---

1. Excluant les dépenses vestimentaires et les dépenses de la section sur les impôtspersonnels, sécurité et dons en argent qui sont collectées au niveau des individus et non au niveau des ménages.

ii) Des variables de revenu

- Il a fallu imputer au moins une variable de revenu pour 7,5 % des individus âgés de 15 ans et plus.

iii) Des variables catégoriques

- Il a fallu imputer au moins une variable catégorique pour 9,8 % des ménages. Pour environ 73 % d'entre eux, seulement une variable a été imputée. À l'échelle provinciale et territoriale, les taux d'imputation varient entre un minimum de 4,3 % pour Terre-Neuve-et-Labrador et un maximum de 23,9 % pour les Territoires du Nord-Ouest. Les variables catégoriques qui sont imputées se trouvent aux sections suivantes du questionnaire: «Caractéristiques du logements», «Équipement associé au logement», «Mode d'occupation du logement», et «Tabac et dépenses diverses», pour les variables se rapportant aux achats par le biais de vente directe.

Introduction

L'Enquête sur les dépenses des ménages (EDM) est une enquête annuelle qui collecte les dépenses et les revenus des ménages à l'aide d'entrevues personnelles. L'échantillon de l'EDM 2009 est composé de 16 758 ménages[2] répartis dans les dix provinces et les trois territoires. Cette taille représente une diminution d'environ 25 % de la taille habituelle (avant 2008) de l'échantillon de l'EDM, ce qui aura un impact à la hausse sur les coefficients de variation des estimations. La collecte est effectuée au cours des mois de janvier, février et mars et les revenus et dépenses sont obtenus pour la période du 1er janvier au 31 décembre de l'année précédente. À la suite d'un remaniement en 1997, cette enquête remplace l'Enquête périodique sur les dépenses des familles et l'Enquête sur l'équipement ménager (avec ajustements aux questionnaires et aux échantillons).

L'EDM, comme toutes les enquêtes, est sujette aux erreurs et ce malgré toutes les précautions prises aux différentes étapes de l'enquête pour les contrôler. Quoiqu'il n'existe pas de mesure exhaustive de la qualité des données d'une enquête, certaines mesures de qualité produites aux différentes étapes de l'enquête peuvent fournir à l'utilisateur de l'information pertinente pour une bonne interprétation des données.

Ce document vise donc à présenter les indicateurs de qualité produits pour l'enquête sur les dépenses des ménages de 2009. On y trouve les indicateurs de qualité usuels généralement utiles aux utilisateurs pour l'interprétation des données tels que les coefficients de variation, les taux de réponse et de non-réponse, les taux de glissement et les taux d'imputation.

Les indicateurs de qualité ont été classés selon les principaux types d'erreur que l'on retrouve dans une enquête. Les erreurs d'échantillonnage, c'est-à-dire les erreurs dues au fait que les inférences tirées au sujet de l'ensemble de la population sont fondées sur des renseignements recueillis auprès d'un échantillon de la population et non pas auprès de toute la population, sont présentées dans la section 1. Les sections suivantes couvrent les erreurs non dues à l'échantillonnage. La non-réponse et les erreurs de couverture sont d'abord discutées dans les sections 2 et 3. Les erreurs de réponse et les erreurs de traitement sont traitées dans les sections 4 et 5 respectivement.

Ce document est axé sur la qualité des données. Pour une description détaillée de la méthodologie de l'enquête, on peut consulter la référence [1].

2. Au départ, l'échantillon compte 20 146 logements desquels on doit identifier et enlever les logements inadmissibles (voir section 2.1) pour en arriver à 16 758 ménages auprès desquels on collecte les dépenses et les revenus.

1. Les erreurs d'échantillonnage

Les erreurs d'échantillonnage découlent du fait que les inférences que l'on tire au sujet de l'ensemble de la population à partir de l'enquête sont fondées sur l'information que l'on a recueillie auprès d'un échantillon de la population et non pas auprès de toute la population. Outre le plan de sondage et la méthode d'estimation qui ont été appliqués pour l'enquête sur les dépenses des ménages, la taille d'échantillon ainsi que la variabilité de chaque caractéristique sont des facteurs déterminants de l'erreur d'échantillonnage. Les caractéristiques qui sont rares ou qui sont distribuées de façon très différente dans la population auront une erreur d'échantillonnage plus grande que les caractéristiques qu'on observe plus fréquemment ou qui sont plus homogènes dans la population.

1.1 Les mesures de l'erreur d'échantillonnage

L'erreur-type est une mesure standard de l'erreur d'échantillonnage. L'erreur-type correspond au degré de variation de l'estimation considérant qu'un échantillon particulier a été choisi, plutôt qu'un autre, parmi tous les échantillons possibles de même taille sous le même plan de sondage. Comme l'EDM utilise un plan d'échantillonnage et une méthode d'estimation complexes, on estime l'erreur-type à partir d'une méthode de rééchantillonnage appelée le bootstrap. Jusqu'à l'année de référence 2002, l'EDM utilisait la méthode de rééchantillonnage du jackknife afin de produire les estimations de l'erreur type. Il a été décidé d'utiliser la méthode du bootstrap à compter de l'EDM de 2003 principalement en raison du fait que la Division de la statistique du revenu (DSR) allait diffuser les médianes des dépenses et requérait donc les coefficients de variation de celles-ci. La méthode du bootstrap permet l'estimation de la variance de statistiques non lisses comme les quantiles. Pour plus de détails sur cette méthode, voir la référence [2].

Le coefficient de variation (CV) est également une mesure de fiabilité de l'estimation fréquemment utilisée. Il exprime tout simplement l'erreur-type en pourcentage de l'estimation. Ainsi, si on obtient une estimation Y pour une certaine caractéristique et que SE correspond à l'erreur-type estimée, le CV sera (SE/Y) x 100.

Finalement, l'erreur-type ou le coefficient de variation peuvent être utilisés pour dériver une autre mesure de précision des estimations, l'intervalle de confiance. Cette mesure indique le niveau de confiance selon lequel la valeur réelle dans la population pour une caractéristique qu'on a observée se trouve à l'intérieur de certaines limites. Un intervalle avec un niveau de confiance de 95 % correspond à l'estimation obtenue à partir de l'échantillon ± 2 erreur-type : (Y ± 2 SE).[3] Cela signifie que si on répétait l'échantillonnage un grand nombre de fois, chaque échantillon fournirait un intervalle différent et 95 % des intervalles contiendraient la vraie valeur de la caractéristique. De façon similaire, si on répétait l'échantillonnage, l'intervalle Y± SE contiendrait la vraie valeur dans 68 % des cas.

3. L'intervalle de confiance se calcule directement à partir du CV de façon similaire, soit Y ± 2 (CV x Y)/100.

1.2 Les coefficients de variation

Des estimations des coefficients de variation sont calculées pour les estimations de plusieurs caractéristiques collectées dans l'EDM. Les CV pour les estimations de la moyenne des dépenses par ménage (pour les catégories de dépenses détaillées), ainsi que pour les caractéristiques des logements et l'équipement ménager, sont disponibles sur demande à l'échelle nationale et provinciale (Division de la statistique du revenu, 1-888-297-7355 : revenu@statcan.gc.ca)

On doit noter que les estimations de CV ne tiennent pas compte du fait que certaines données ont été imputées; par conséquent, les CV présentés peuvent sous-estimer les valeurs réelles. Pour la plupart des variables, le taux d'imputation est peu élevé (voir section 5) et les CV fournis représentent une bonne estimation des vrais CV. Cependant, il est important de tenir compte à la fois du CV et du taux d'imputation si on s'intéresse à la fiabilité de dépenses détaillées ayant un taux d'imputation élevé.

Le tableau 1.1 donne un aperçu des CV des estimations des moyennes par ménage à l'échelle nationale, provinciale et territoriale pour l'estimation de quelques-unes des catégories de dépenses sommaires et pour le revenu.

Tableau 1.1
Coefficients de variation par province et territoire ainsi qu'à l'échelle nationale pour l'estimation de la moyenne des dépenses par ménage pour quelques catégories de dépenses sommaires et pour l'estimation du revenu moyen

	Can.	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Yn	T.N.-O.	Nt
	%													
Dépenses totales	1,0	1,4	1,9	1,8	2,2	1,5	2,0	2,1	1,7	1,7	2,3	5,1	7,6	6,0
Dépenses de consommation	0,8	1,5	1,8	1,6	2,1	1,3	1,4	1,8	1,7	1,8	2,4	4,6	6,8	5,1
Alimentation	0,7	1,4	2,0	1,5	1,8	1,4	1,4	2,2	1,6	2,0	1,9	3,7	4,8	4,3
Logement	1,0	2,3	2,8	2,3	2,2	2,2	1,8	2,8	2,0	2,4	2,9	6,1	6,2	8,3
Entretien ménager	1,3	2,4	2,9	2,3	4,6	2,3	2,7	3,2	2,5	2,5	2,6	5,5	5,7	9,4
Ameublement	2,3	4,6	6,5	4,7	4,4	4,7	4,2	5,0	5,2	5,6	7,0	10,6	13,3	15,9
Habillement	1,7	2,9	3,9	3,4	3,2	2,9	3,0	3,4	3,3	3,6	5,5	7,7	6,8	6,0
Transport	1,6	3,7	4,3	4,1	3,6	3,5	2,8	4,3	4,1	4,2	4,1	7,9	12,2	13,1
Soins de santé	1,8	4,4	4,8	3,2	4,1	3,0	4,2	3,9	3,5	3,9	4,3	7,9	9,9	20,6
Soins personnels	1,4	2,8	3,6	2,5	2,6	2,5	2,6	3,3	3,1	3,3	3,0	6,1	7,6	7,2
Loisirs	2,8	4,8	5,9	4,4	12,9	3,6	3,0	3,7	4,5	4,6	15,2	10,1	17,4	7,4
Lecture et autres imprimés	2,2	4,9	7,6	4,2	4,4	3,9	4,7	4,4	5,0	4,9	4,1	11,9	14,9	16,1
Éducation	4,3	14,6	13,8	10,9	11,0	8,0	8,2	10,5	10,4	8,1	9,7	18,3	30,1	60,2
Tabac, boissons alcoolisées	2,3	5,0	7,0	5,0	4,3	3,9	5,2	5,2	5,4	5,5	4,7	9,7	10,5	9,1
Jeux de hasard (net)	5,7	6,8	15,9	8,3	7,7	9,9	6,7	9,9	22,3	28,6	9,0	15,0	18,2	10,8
Dépenses diverses	5,6	6,6	9,2	10,2	20,2	4,8	12,8	6,4	6,7	6,9	10,2	12,1	12,1	28,7
Impôts personnels	2,3	2,9	5,0	4,3	3,5	3,8	4,9	5,9	3,9	3,8	4,4	8,1	11,3	10,0
Assurances individuelles et cotisations de retraite	6,2	2,7	3,6	2,4	2,8	3,2	14,8	2,4	3,5	2,7	2,7	9,9	11,0	8,9
Dons et contributions	6,9	7,3	14,1	9,7	8,1	20,0	13,0	9,8	9,3	10,0	13,8	14,4	19,0	22,7
Revenu	0,9	1,2	1,9	1,7	1,5	1,6	2,0	2,1	1,6	1,8	2,2	5,6	8,1	6,2

Le tableau 1.2 donne un aperçu des CV pour les caractéristiques des logements et de l'équipement ménager à l'échelle nationale, provinciale et territoriale pour l'estimation de quelques-unes des catégories.

Tableau 1.2
Coefficients de variation par province et territoire ainsi qu'à l'échelle nationale pour quelques catégories des caractéristiques des logements et l'équipement ménager

	Can.	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Yn	T.N.-O.	Nt
	%													
Propriétaire	1,0	2,0	2,8	2,3	1,9	2,6	1,9	2,9	2,3	2,0	2,6	11,6	10,8	18,5
Locataire	1,9	7,1	6,6	4,6	6,2	3,6	3,8	5,7	6,4	4,9	4,5	16,2	9,1	4,5
Machine à laver	0,8	1,0	1,9	1,7	1,1	1,2	1,7	2,5	1,8	1,5	2,2	5,4	5,1	2,6
Sécheuse	0,8	1,1	1,8	1,7	1,2	1,4	1,7	2,4	1,6	1,3	2,1	4,9	4,8	2,7
Lave-vaisselle	1,2	3,9	3,3	3,9	3,2	2,6	2,3	3,3	3,4	1,8	2,8	12,8	11,2	16,2
Congélateur	1,2	1,8	2,8	2,4	2,0	2,8	2,7	2,9	2,0	2,5	2,9	6,7	5,4	7,9
Four à micro-ondes	0,3	0,6	0,9	0,7	0,8	0,8	0,6	0,8	1,1	0,7	0,9	3,1	1,8	3,4
Téléphone cellulaire	0,7	1,9	2,5	2,1	2,0	2,0	1,2	2,1	1,6	1,2	1,3	6,4	7,3	22,7
Lecteur de disques compacts	0,8	1,5	2,5	1,6	1,5	1,7	1,5	2,3	1,6	1,8	1,7	3,6	4,1	4,5
Télédistribution	1,2	2,4	5,8	2,8	3,2	2,4	2,5	3,0	3,8	3,0	2,1	8,5	8,3	12,6
Antenne parabolique	2,9	5,6	7,4	6,0	3,9	6,7	5,7	6,7	5,9	5,9	9,0	13,1	8,2	15,2
Lecteur DVD	0,6	1,3	2,0	1,3	1,4	1,2	1,1	1,3	1,3	1,2	1,1	2,7	3,0	3,5
Ordinateur personnel	0,7	1,9	2,3	1,6	1,9	1,5	1,2	1,8	1,5	1,3	1,3	5,0	4,3	6,2
Connexion tél. régulière à un ordinateur (Modem)	7,3	25,6	31,5	18,2	17,2	11,5	13,0	16,1	13,6	22,5	24,7	46,6	55,9	55,2
Connexion tél. haute vitesse à un ordinateur	2,5	4,7	6,4	4,3	4,0	6,7	4,9	5,2	3,8	6,0	6,4	9,9	16,2	16,4
Connexion câble à un ordinateur	2,3	7,9	15,4	7,7	9,6	4,1	4,9	5,9	8,9	5,5	3,9	15,8	12,1	35,8
Utilisation d'Internet (maison)	0,8	2,1	3,0	1,8	2,2	1,7	1,5	1,9	1,8	1,5	1,3	6,4	4,6	9,4
Véhicules possédés (un)	1,8	3,6	5,5	4,2	3,2	3,6	3,7	3,9	6,1	4,3	3,8	9,7	9,8	14,7
Véhicules possédés (2 ou plus)	1,7	4,5	4,4	4,0	2,9	4,0	3,6	3,9	4,3	3,0	3,6	7,7	15,9	26,7

1.3 Modèle pour dériver une approximation du CV

Des estimations pour différents domaines d'intérêt (par ex., par quintile de revenu) sont disponibles pour les catégories de dépenses sommaires dans la publication *Les habitudes de dépenses au Canada* (référence [4]). Des estimations pour différents domaines d'intérêt pour les catégories de dépenses détaillées sont également disponibles sur demande auprès de la Division de la statistique du revenu. (Pour plus de détails sur les tableaux disponibles auprès de la Division de la statistique du revenu, voir les références [3] ou [4].) Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour tous les différents niveaux d'agrégation qui pourraient intéresser les utilisateurs.

1.3.1 Approximation du CV pour les estimations des domaines

Il est toutefois possible de calculer une approximation du CV en utilisant une relation entre le nombre de ménages de l'échantillon qui ont déclaré des

dépenses pour une catégorie et le CV à un niveau agrégé. Cette relation, basée sur la tendance du CV à croître proportionnellement à une diminution de la racine carrée du nombre de ménages déclarant une dépense, est illustrée ci-après.

Formule d'approximation du CV pour un domaine (un sous-groupe de la population)

Si CV(Y) représente le CV pour l'estimation de la moyenne par ménage d'une certaine caractéristique pour toute la population, alors on peut calculer une approximation du CV de l'estimation de cette caractéristique pour un domaine (que l'on peut considérer comme un sous-groupe de la population tel qu'un type de ménage, un quintile de revenu, un niveau d'urbanisation) à partir de l'équation suivante :

$$CV\ (Y_d) = CV\ (Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

où

n : le nombre de ménages dans l'échantillon
P : l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans la population
n_d : le nombre de ménages de l'échantillon dans le domaine d
P_d : l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans le domaine d

Généralement on utilise le CV, la taille *n* et la proportion *P* à l'échelle nationale pour calculer les approximations pour les différents domaines. Dans le cas où on cherche à calculer une approximation du CV pour un domaine entièrement contenu dans une seule province (par ex. région métropolitaine), il est préférable d'utiliser ces valeurs à l'échelle provinciale puisque les CV provinciaux sont publiés pour l'EDM de 2009 (référence [3]). Il est important de se rappeler que la valeur du CV obtenue par cette approche n'est qu'une approximation du CV.

1.3.2 Approximation du CV à partir du fichier de microdonnées

Les utilisateurs du fichier de microdonnées peuvent se servir d'une autre approche pour dériver une approximation du CV des estimations, qui sera généralement plus performante que celle présentée dans la section précédente pour les CV des catégories de dépenses détaillées. Cette approche est décrite en détails dans la documentation accompagnant le fichier de microdonnées de 2009. Elle est utilisable seulement à partir du fichier de microdonnées puisqu'il est nécessaire d'avoir les données et les poids de chaque ménage pour calculer cette approximation.

Le document sur la qualité des données de l'enquête de 1997 renferme des résultats de l'évaluation de la performance de ces deux méthodes d'approximation du CV.

1.4 Suppression des données non fiables dans les tableaux d'estimations

Comme le coefficient de variation est un indicateur de la fiabilité des données, on aimerait l'utiliser pour déterminer si les estimations devraient être publiées ou non. On considère que les estimations dont le CV est estimé à plus de 33 % ne sont pas suffisamment fiables pour être publiées. Toutefois les estimations des CV ne sont pas calculées pour une grande partie des estimations publiées, la règle de suppression pour les estimations de dépenses est donc basée sur le nombre de ménages qui déclarent une valeur supérieure à zéro.[4]

On peut démontrer que les CV sont en général inférieurs à 33 % si le nombre de ménages qui déclarent une dépense est supérieur à 30. Comme il s'agit d'une règle approximative, certaines estimations peuvent être publiées même si le CV est supérieur à 33 % et certaines estimations ne seront pas publiées malgré un CV inférieur à 33 %. Le document sur la qualité des données de l'EDM 1997 donne les résultats de l'évaluation des risques d'erreurs de l'utilisation de la règle de suppression.

2. La non-réponse

Les erreurs dues à la non-réponse découlent du fait que certains répondants potentiels ne fournissent pas l'information nécessaire ou que cette information s'avère inutilisable. Lorsque le répondant a omis de répondre à certaines questions seulement, on parle alors de non-réponse partielle. Dans ce cas, les données manquantes sont imputées. Les erreurs liées à l'imputation sont présentées dans la section 5 portant sur les erreurs de traitement. Dans la présente section, la non-réponse englobe la non-réponse à la collecte, due principalement à l'impossibilité de contacter le ménage ou au refus des membres du ménage de participer à l'enquête, que ce soit partiellement ou complètement, ainsi que les ménages pour lesquels les données ont été collectées mais sont inutilisables.

Le principal impact de la non-réponse sur la qualité des données est qu'elle peut induire un biais dans les estimations si les caractéristiques des répondants et des non-répondants diffèrent et que cette différence a un impact sur les caractéristiques étudiées. Les taux de non-réponse peuvent être calculés facilement mais ils n'ont qu'une valeur indicative à l'égard de la qualité des données car ils ne permettent pas de mesurer l'importance du biais associé aux estimations. L'ampleur de la non-réponse peut être considérée comme une évaluation des risques de biais dans les estimations.

4. En pratique, on utilise l'estimation de la proportion des ménages déclarant une dépense que l'on multiplie par la taille d'échantillon.

2.1 Le taux de réponse, les taux de non-réponse et les taux de vacance

Dans l'EDM, comme les unités sélectionnées sont des logements, les intervieweurs doivent d'abord identifier les logements inadmissibles, c'est-à-dire les logements occupés par des personnes ne faisant pas partie de la population cible, les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise) et les logements vacants (inoccupés, saisonniers ou en construction).

Parmi les logements admissibles, on évalue ensuite la proportion des ménages qui n'ont pas répondu à l'enquête, qu'on appelle le taux de non-réponse à la collecte. Ces derniers comprennent les ménages qui ont refusé de participer à l'enquête et les ménages où il a été impossible d'établir un contact avec les répondants parce qu'ils étaient absents ou encore à cause de circonstances spéciales (problème de langue, maladie ou décès).

Toujours parmi les logements admissibles, on détermine également le taux de données inutilisables. Les données inutilisables correspondent au nombre de ménages pour lesquels le questionnaire était au moins partiellement complet mais qui ont été rejetés lors du traitement des données. Lorsqu'une partie importante des questions sur le revenu ou des questions sur les dépenses a été laissée sans réponse, le questionnaire est classé incomplet et n'est pas utilisé.

Il est à noter que tous les taux de la présente section sont non pondérés. Pour l'Enquête sur les dépenses des ménages de 2009, le taux de réponse final est de 64,5 %. Le tableau 2.1-1 présente le taux de réponse final ainsi que la taille de l'échantillon (ménages admissibles) ventilée selon les refus, les unités non contactées, les données inutilisables et les données utilisables. Ces informations sont fournies à l'échelle nationale, provinciale et territoriale.

Tableau 2.1-1
Taille de l'échantillon et taux de réponse par province et territoire ainsi qu'à l'échelle nationale

	Ménages admissibles	Non contactés	Ayant refusé de participer	Inutilisables	Utilisables	Taux de réponse final (à l'estimation)[1]
	Nombre					%
Canada	**16 758**	**1 807**	**3 644**	**496**	**10 811**	**64,5**
Terre-Neuve-et-Labrador	1 347	113	207	19	1 008	74,8
Île-du-Prince-Édouard	789	53	212	9	515	65,3
Nouvelle-Écosse	1 515	167	369	51	928	61,3
Nouveau-Brunswick	1 377	117	290	17	953	69,2
Québec	2 040	167	505	63	1 305	64,0
Ontario	2 582	371	579	93	1 539	59,6
Manitoba	1 413	182	354	25	852	60,3
Saskatchewan	1 345	114	249	26	956	71,1
Alberta	1 572	196	336	89	951	60,5
Colombie-Britannique	1 790	186	406	83	1 115	62,3
Yukon	372	31	65	4	272	73,1
Territoires du Nord-Ouest	376	82	53	11	230	61,2
Nunavut	240	28	19	6	187	77,9

1. Utilisables/ménages admissibles x 100

Le tableau 2.1-2 présente le taux de non-réponse final, le taux de non-réponse à la collecte, ventilé selon les refus et les unités non contactées ainsi que le taux de données inutilisables, en raison des questionnaires incomplets. On y trouve également le taux de vacance. Ces taux sont fournis à l'échelle nationale, provinciale et territoriale.

Notons que les taux de vacance présentés aux tableaux de la section 2 incluent les logements vacants (inoccupés, saisonniers ou en construction) ainsi que les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise).

Tableau 2.1-2
Taux de non-réponse et taux de vacance par province et territoire ainsi qu'à l'échelle nationale

	Taux de vacance	Taux de non-réponse à la collecte			Taux de données inutilisables			Taux de non-réponse final (à l'estimation)
		Total	Pas de contact	Refus	Total	Incomplet	Non équilibré	
	%							
Canada	**15,0**	**32,5**	**10,8**	**21,7**	**3,0**	**0,7**	**2,3**	**35,5**
Terre-Neuve-et-Labrador	15,0	23,8	8,4	15,4	1,4	0,3	1,1	25,2
Île-du-Prince-Édouard	19,3	33,6	6,7	26,9	1,1	0,4	0,8	34,7
Nouvelle-Écosse	18,4	35,4	11,0	24,4	3,4	0,8	2,6	38,7
Nouveau-Brunswick	16,9	29,6	8,5	21,1	1,2	0,1	1,2	30,8
Québec	12,2	32,9	8,2	24,8	3,1	0,9	2,2	36,0
Ontario	11,5	36,8	14,4	22,4	3,6	0,6	3,0	40,4
Manitoba	11,5	37,9	12,9	25,1	1,8	0,8	1,0	39,7
Saskatchewan	15,3	27,0	8,5	18,5	1,9	0,1	1,8	28,9
Alberta	14,4	33,8	12,5	21,4	5,7	1,2	4,5	39,5
Colombie-Britannique	16,6	33,1	10,4	22,7	4,6	0,8	3,8	37,7
Yukon	17,7	25,8	8,3	17,5	1,1	0,5	0,5	26,9
Territoires du Nord-Ouest	19,8	35,9	21,8	14,1	2,9	1,3	1,6	38,8
Nunavut	20,7	19,6	11,7	7,9	2,5	1,7	0,8	22,1

2.2 La non-réponse selon le niveau d'urbanisation

La non-réponse varie selon le niveau d'urbanisation. Les divers taux à l'échelle nationale sont présentés par niveau d'urbanisation dans le tableau 2.2.[5]

Tableau 2.2
Taux de non-réponse et taux de vacance par niveau d'urbanisation

	Taux de vacance	Taux de non-réponse à la collecte			Taux de données Inutilisables			Taux de non-réponse final (à l'estimation)
		Total	Pas de contact	Refus	Total	Incomplet	Non équilibré	
Centres de population	%							
1 000 000 ou plus	10,9	34,4	11,6	22,8	3,8	0,8	3,1	38,3
500 000 à 999 999	10,0	36,5	13,1	23,4	3,6	1,0	2,6	40,1
250 000 à 499 999	8,8	38,2	12,4	25,8	4,0	0,6	3,3	42,2
100 000 à 249 999	10,8	33,9	11,4	22,5	2,0	0,4	1,6	35,8
30 000 à 99 999	11,7	32,1	8,5	23,6	2,7	0,7	2,0	34,8
1 000 à 29 999	15,8	28,8	10,9	17,9	2,6	0,6	2,0	31,4
Rural	28,8	27,1	7,6	19,4	2,4	0,5	1,9	29,5
Total	15,0	32,5	10,8	21,7	3,0	0,7	2,3	35,5

5. Des tableaux portant sur les taux de non-réponse selon le niveau d'urbanisation par province sont disponibles sur demande auprès de la Division des méthodes d'enquêtes auprès des ménages.

2.3 La non-réponse selon les strates de revenu

Il est impossible de comparer le taux de réponse selon le revenu car cette information n'est pas accessible pour les non-répondants. Toutefois, le plan d'échantillonnage de l'EPA, utilisé pour l'EDM, a été conçu de sorte à former, dans les régions métropolitaines de recensement, des strates d'aires géographiques présentant une plus forte concentration de ménages à revenu élevé. Bien que le nombre de strates à revenu élevé demeure relativement petit (51 sur un total de 1060 strates), la comparaison des taux de non-réponse dans ce groupe par rapport à l'ensemble des autres strates fournit de l'information pertinente sur l'effet potentiel de la non-réponse (voir le tableau 2.3).

Mentionnons que l'ensemble des autres strates comprend, en plus des strates régulières, les quatre types de strates suivants : les strates à taux de vacance élevé, les strates à coût élevé, les strates ayant une concentration d'immigrants et les strates ayant une concentration d'autochtones. La portion de l'échantillon de l'EDM allouée pour ces quatre derniers types de strates étant plus faible, les résultats de ceux-ci ne sont pas présentés individuellement au tableau 2.3.

Tableau 2.3
Comparaison des taux de non-réponse et de vacance des strates à revenu élevé par rapport aux autres strates

	Taux de vacance	Taux de non-réponse à la collecte			Taux de données inutilisables			Taux de non-réponse final (à l'estimation)
		Total	Pas de contact	Refus	Total	Incomplet	Non équilibré	
	%							
Revenu élevé	8,5	37,9	13,1	24,7	3,1	1,1	2,0	40,9
Autres	15,5	32,0	10,6	21,5	2,9	0,6	2,3	35,0
Total	15,0	32,5	10,8	21,7	3,0	0,7	2,3	35,5

2.4 L'ajustement pour la non-réponse

Pour compenser la non-réponse, les poids de l'EDM sont gonflés par l'inverse du taux de réponse pondéré à l'intérieur de certains groupes prédéfinis. À la suite du remaniement du plan de sondage de l'EPA, les groupes d'ajustement de non-réponse ont été remaniés. Tout comme pour les années précédentes, ces groupes sont définis en fonction des différents niveaux d'urbanisation dans chaque province et de régions géographiques infra-provinciales pour les provinces du Québec, de l'Ontario et de la Colombie-Britannique. De plus, des groupes spécifiques d'ajustement de non-réponse sont créés pour les strates à revenu élevé. Comme on l'a vu à la section précédente, le nombre de strates à revenu élevé a augmenté avec l'introduction du nouveau plan de sondage de l'EPA. Ainsi il est dorénavant possible de former dans toutes les provinces, à l'exception de l'Île-du-Prince-Édouard, des groupes d'ajustement de non-réponse pour les strates à revenu élevé. On ne peut former un tel groupe à l'Île-du-Prince-Édouard, puisque cette province ne possède pas de strates à revenu élevé.

Les taux pondérés diffèrent des taux présentés dans cette section puisqu'ils tiennent compte du poids de sondage de chaque ménage. Une description algébrique de l'ajustement pour la non-réponse est présentée à l'annexe A.

L'ajustement des poids pour la non-réponse permet de tenir compte des différences au niveau de la non-réponse par niveau d'urbanisation (tel qu'illustré dans la section 2.2) et régions géographiques ou par groupes de strates à revenu élevé. Il permettra de réduire le biais dans la mesure où les caractéristiques des répondants et des non-répondants sont similaires pour un même niveau d'urbanisation et région géographique ou encore pour un même groupe de strates à revenu élevé.

Il est à noter qu'un groupe d'ajustement de non-réponse peut être combiné à un autre groupe si le nombre de ménages dans le groupe est trop petit ou si le facteur d'ajustement est trop élevé.

3. Les erreurs de couverture

Lors de la conception de l'enquête, on a défini la population visée, qu'on appelle la population cible. Rappelons d'abord la définition de cette population cible pour l'EDM puisqu'une bonne compréhension de la population cible est nécessaire à une bonne interprétation des données de l'enquête. Il est important de préciser que l'EDM utilise la base de sondage de l'enquête sur la population active (EPA).

La population cible

> La population cible correspond aux individus vivant dans les ménages privés. Les pensionnaires d'établissement institutionnel tel que les prisons, les hôpitaux pour malades chroniques, les résidences pour personnes âgées ainsi que les membres d'ordres religieux et d'autres groupes vivant en communauté, les membres des Forces Armées vivant dans les camps militaires et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres sont donc exclus. On exclut aussi les représentants officiels de pays étrangers qui vivent au Canada et leurs familles ainsi que les résidents des réserves indiennes et des terres publiques (à l'exception des Territoires). L'enquête couvre donc près de 98 % de la population dans les 10 provinces. Au Yukon, la couverture de la population est de 92 %, celle des Territoires du Nord-Ouest correspond à 92 % et la couverture du Nunavut est de 91 %[6].

6. Pour ce qui est de la proportion de ménages, la couverture du Yukon, des Territoires du Nord-Ouest et du Nunavut correspond respectivement à 92 %, 93 % et 92 % des ménages.

Nous n'avons pas recueilli de renseignements auprès des personnes qui vivent temporairement loin de leur famille (par exemple, les étudiants d'université) parce que nous obtenons l'information auprès de leur famille, si elle est choisie dans l'échantillon.

Les erreurs de couverture proviennent d'une représentation inadéquate de la population cible à partir des unités de la base de sondage. Certaines unités de la population cible peuvent être omises de la base de sondage, on parle alors d'un sous-dénombrement. D'autres unités qui ne sont pas dans la population cible peuvent être incluses par erreur ou certaines unités peuvent être incluses plus d'une fois, ces unités sont responsables du surdénombrement.

3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement

Dans l'EDM, la sélection de l'échantillon est effectuée en utilisant une liste de logements dans chaque grappe sélectionnée. L'omission de logements lors de la création de la liste, les nouveaux logements qui s'ajoutent entre la création de la liste et la visite des intervieweurs (principalement dans les secteurs en développement) ainsi que la classification erronée de logements vacants contribuent au sous-dénombrement. Le fait d'inclure des logements qui ne seraient pas dans les limites de la grappe est une source de surdénombrement. De façon similaire, des erreurs peuvent se glisser lors de la collecte des données, lors de l'identification des personnes qui sont membres du ménage sélectionné. Ces erreurs contribuent également au sous-dénombrement et au surdénombrement.

De plus, tel que décrit à la section 2.4, des méthodes de repondération sont mises en place pour tenir compte de la non-réponse. Lors de ces ajustements, il est cependant impossible de corriger les poids d'enquête pour faire en sorte que tous les sous-groupes de la population soient bien représentés.

Une bonne représentation de la population cible est indispensable à la production d'estimations de dépenses réalistes. Il est nécessaire que l'échantillon représente adéquatement les individus de la population cible et la distribution des ménages selon leur taille.

On observe généralement un sous-dénombrement net du nombre de personnes et du nombre de ménages dans l'EDM, que l'on rectifie par un ajustement des poids en utilisant des données auxiliaires, ou repères, qui sont basées sur des estimations démographiques post-censitaires. Le taux de glissement (voir annexe A) est une mesure du pourcentage d'écart entre les estimations provenant de ces données auxiliaires et les estimations de l'enquête calculées en utilisant des poids non ajustés avec ces données.[7] Le glissement représente donc l'effet cumulé du sous-dénombrement et du débalancement dans certains sous-groupes de la population créé par la non-réponse à l'enquête qui n'a pu être corrigée à l'étape de la repondération.

7. On utilise les sous-poids, c'est-à-dire les poids de sondage ajustés pour la non-réponse (voir Annexe A).

Les taux de glissement par groupe d'âge à l'échelle nationale, provinciale et territoriale sont présentés dans le tableau 3.1 alors que les taux de glissement selon les tailles de ménages que l'on utilise lors de l'ajustement des poids se trouvent dans le tableau 3.2. Un taux positif correspond à un surdénombrement du nombre de personnes ou de ménages dans l'enquête.

Tableau 3.1
Taux de glissement pour les provinces et les territoires selon le groupe d'âge

Âge	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Canada (sans les territoires)	Yukon	Territoires du Nord-Ouest	Nunavut
	%													
0-6	-10,1	5,2	-0,1	6,7	-4,8	-24,9	-11,2	-15,6	-13,7	-11,7	-14,8	-33,5	-24,0	-11,7
7-17	-9,4	2,7	-13,8	-4,7	-19,2	-16,7	-9,4	-12,6	-11,3	-6,5	-14,4			
18-24	-26,9	-33,4	-33,3	-14,6	-5,8	-23,8	-26,8	-18,7	-26,8	-40,1	-22,5			
25-34	-28,0	-4,2	-14,0	-16,7	-13,3	-33,7	-10,4	-7,2	-27,4	-28,6	-25,0			
35-44	-10,6	-8,0	-13,3	-12,3	-14,3	-23,0	-18,5	-11,8	-16,3	-18,0	-18,5	-31,6	-26,1	-24,3
45-54	-15,4	-17,9	-20,3	-6,9	-24,1	-16,3	-12,4	-6,5	-18,3	-23,2	-18,8			
55-64	0,1	-20,1	-21,2	-9,0	-1,8	-7,8	2,6	-14,3	-20,6	-17,4	-9,0			
65 +	-7,8	17,6	2,0	-1,3	-0,7	-8,4	-1,3	-6,2	-8,3	-9,5	-5,7			
Total	-12,7	-7,3	-14,4	-7,9	-11,2	-19,0	-10,5	-11,0	-18,3	-19,3	-16,1	-32,0	-25,5	-19,4

Tableau 3.2
Taux de glissement pour les provinces et territoires selon la taille de ménage

	Ménages	Ménages d'une personne	Ménages de deux personnes	Ménages de trois personnes et plus
	%			
Canada	-12,2	-13,6	-4,7	-18,2
Terre-Neuve-et-Labrador	-7,0	-4,8	-1,7	-14,5
Île-du-Prince-Édouard	-5,1	-0,4	-6,6	-6,6
Nouvelle-Écosse	-10,3	-6,2	-6,9	-18,1
Nouveau-Brunswick	-9,2	-20,7	-4,5	-5,7
Québec	-7,6	-8,5	-1,1	-13,8
Ontario	-13,3	-15,4	-1,5	-21,3
Manitoba	-9,4	-14,6	-4,2	-10,0
Saskatchewan	-6,3	-5,2	-0,4	-13,6
Alberta	-17,5	-17,1	-15,4	-19,6
Colombie-Britannique	-17,7	-21,0	-11,6	-20,6
Yukon	-27,1	-18,8	-28,0	-34,2
Territoires du Nord-Ouest	-29,8	-32,2	-42,8	-14,8
Nunavut	-18,3	-33,6	-47,7	3,7

3.2 L'ajustement au niveau de la population et des ménages

Afin de corriger le problème de représentativité illustré dans le tableau 3.1 et réduire le biais qui en découle, on ajuste les données de l'enquête lors de la pondération en utilisant les estimations démographiques pour les groupes d'âge définis dans ces tableaux, et ce pour chaque province et territoire. Pour plus de détails sur la méthodologie de l'ajustement, voir les références [1] et [5]. Cet ajustement réduit le biais mais ne l'élimine pas complètement si les caractéristiques des individus que l'on a omis de l'enquête diffèrent de celles des individus inclus pour un même groupe d'âge dans une province ou un territoire.

On notera également que l'efficacité de l'ajustement à partir des estimations démographiques dépend beaucoup de la qualité de ces estimations démographiques et de leur exactitude à bien représenter la population cible de l'enquête. Les estimations démographiques ne sont pas exemptes d'erreurs. Ce sont des estimations post-censitaires basées sur les comptes de population du Recensement de 2001, ajustées pour le sous-dénombrement net et qui tiennent compte des récentes statistiques sur la migration, les naissances, la mortalité, etc. Ces estimations démographiques sont ajustées pour tenir compte de certaines exclusions spécifiques aux enquêtes auprès des ménages comme les personnes vivant en institutions. Conceptuellement, elles diffèrent légèrement de la population cible de l'EDM en incluant les personnes habitant les ménages collectifs qui ne sont pas des institutions, par exemple les membres de groupes vivant en communauté et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres. Cependant, on considère cette différence négligeable puisque ces individus représentent moins de 0,4 % de la population canadienne.

Pour corriger le problème de représentativité de l'échantillon par rapport au nombre de ménages selon leur taille, illustré dans le tableau 3.2, on ajuste les données de l'enquête à l'aide de données auxiliaires. En ajustant les poids de l'EDM pour refléter les estimations post-censitaires du nombre de ménages selon la taille, on vise à compenser le biais qui découle d'une représentation inadéquate des ménages. Toutefois on n'éliminera pas nécessairement ce biais si les caractéristiques des ménages non interviewés (omis ou non-répondants) diffèrent de celles des ménages répondants pour une même taille de ménage. Comme dans le cas des estimations démographiques de population, l'efficacité de l'ajustement dépend de la qualité des données auxiliaires sur le nombre de ménages.

En plus des estimations démographiques quant aux groupes d'âge par province, deux autres ensembles de données auxiliaires sont utilisés pour ajuster les données d'enquêtes lors de la pondération en vue d'en améliorer la représentativité. Le premier ensemble de données est utilisé pour contrôler le nombre d'enfants et d'adultes dans certaines grandes villes. Pour le second, des comptes par grandes catégories de revenu en salaires et traitements sont utilisés lors de l'ajustement des poids afin d'assurer une certaine cohérence entre les distributions de revenu de l'EDM et celles provenant de sources externes.

4. Les erreurs de réponse

Les erreurs de réponse correspondent au manque d'exactitude des réponses aux questions. Elles peuvent être attribuables à différents facteurs parmi lesquels se retrouvent un questionnaire qui nécessite des améliorations, une mauvaise interprétation des questions par les intervieweurs ou les répondants ainsi que des erreurs dans les déclarations des répondants.

Dans l'EDM, les erreurs dans les déclarations des répondants peuvent survenir pour différentes raisons. Premièrement, il y a les erreurs de rappel qui se produisent lorsque le répondant oublie les dépenses effectuées durant la période de l'enquête (qui correspond à l'année civile) ou fournit une valeur erronée à cause de l'intervalle de temps écoulé entre le moment de l'achat et la date de l'entrevue. Les erreurs de rappel sont probablement la source d'erreur de réponse la plus importante de l'enquête étant donné que la période de référence est longue (12 mois) et qu'une grande variété d'information est demandée.

Pour diminuer l'ampleur de ce type d'erreur, on encourage le répondant à consulter divers documents (factures, relevés bancaires,...) pour fournir des données plus exactes. Pour déterminer les dépenses pour de petits articles achetés à intervalles réguliers, les intervieweurs suggèrent généralement aux répondants d'estimer la fréquence des achats et le prix généralement payé pour dériver les dépenses sur une période de 12 mois.

Une seconde source d'erreur dans la déclaration des répondants est l'erreur télescopique qui consiste à inclure dans la période de référence des événements qui se sont produits avant ou après celle-ci. Dans l'EDM, on considère que le fait d'utiliser l'année civile est un bon repère pour le début de la période de référence. De plus, comme la période de référence est longue, l'impact de l'erreur télescopique est moins important.

Les réponses par procuration peuvent également contribuer à l'erreur de réponse. Le membre du ménage qui a effectué une dépense est généralement la personne qui peut la rapporter avec la meilleure précision. C'est sûrement le cas par exemple pour les achats personnels. Les dépenses rapportées par personne interposée sont plus susceptibles d'être entachées d'une erreur de réponse et ce type d'erreur risque d'affecter plus sérieusement certaines catégories de dépenses.

Parmi d'autres sources d'erreur de réponse, il ne faut pas oublier le degré de coopération du répondant. Il se peut que le répondant décide pour des raisons personnelles de ne pas mentionner certaines dépenses ou de déformer les faits.

Dans l'EDM, on considère également que le fardeau imposé au répondant, étant donné la longueur de l'entrevue et la grande variété d'articles à rapporter, ainsi que le rythme de l'entrevue peuvent entraîner la fatigue du répondant et avoir un impact sur la qualité des réponses obtenues. Le temps d'entrevue varie beaucoup d'un ménage à l'autre selon la taille du ménage, le revenu et plusieurs autres caractéristiques.

Quoique l'on considère que les erreurs de réponses sont une source importante d'erreur dans une entrevue rétrospective, elles demeurent l'aspect lié à la qualité des données le plus difficile à mesurer. Il faut généralement entreprendre des études spéciales très coûteuses pour tenter de les mesurer. On tente de les réduire en implantant des techniques de sondage développées à cet effet.

5. Les erreurs de traitement

Des erreurs peuvent se produire lors de toutes les manipulations des données. Les principales étapes du traitement des données sont le codage des réponses, la saisie des données, la vérification, l'imputation des non-réponses partielles et la pondération. Dans l'EDM, différentes procédures sont appliquées à chacune des étapes afin de réduire au minimum les erreurs de traitement et les estimations de l'enquête sont comparées avec d'autres sources avant d'être publiées. Les erreurs qui peuvent découler des ajustements effectués lors de la pondération ont été décrites dans les sections 2 et 3. La présente section couvre principalement les autres types d'erreurs de traitement.

En raison du passage à un mode de collecte assisté par ordinateur en 2006, les mesures de traitement des données et de contrôle de la qualité ont été modifiées. Des vérifications automatiques intégrées à ce questionnaire ont remplacé les contrôles d'équilibre et les vérifications dans les bureaux régionaux qui se faisaient auparavant. Pour l'EDM de 2009, les intervieweurs ont saisi les réponses sur ordinateur portatif et procédé à une première vérification simultanément. Ainsi, les contrôles d'intervalle, qui présentaient des minimums et des maximums pour certains achats ont été appliqués si l'intervieweur entrait un montant inusité. D'autres vérifications visaient les réponses incohérentes si, par exemple, le ménage occupait son logement en location, mais qu'aucun loyer n'était payé.

Le traitement des données de l'EDM comprend également l'imputation de données pour la non-réponse partielle. Il y a non-réponse partielle lorsque le répondant refuse de répondre ou ne connaît pas la réponse à certaines questions particulières. L'approche pour imputer les données diffère selon qu'il s'agit de données catégoriques ou continues. Les données catégoriques peuvent prendre uniquement quelques valeurs (comme les questions pour lesquelles la réponse est oui ou non et les questions sur le type de logement habité) alors que les données continues peuvent prendre n'importe quelle valeur numérique (comme les revenus et les dépenses).

Les données sur le revenu et les dépenses sont imputées au moyen de la technique du plus proche voisin. L'imputation se fait sur un groupe de variables à la fois, ce groupe étant formé de sorte à tenir compte des relations entre ces variables. Un groupe correspond généralement à une section du questionnaire. Pour chaque groupe, les valeurs manquantes d'un receveur (un ménage qui a des données manquantes pour au moins une de ces variables) sont imputées à partir des données du plus semblable parmi tous les donneurs (les ménages qui n'ont aucune donnée manquante pour ces variables). Pour chaque receveur, on détermine le donneur le plus semblable comme étant celui qui a la valeur

minimale pour une certaine fonction de distance. Cette fonction est basée sur des variables d'appariement choisies parce qu'elles sont corrélées avec les variables à imputer. Par exemple, le revenu total du ménage fait partie des variables d'appariement pour presque toutes les sections portant sur les dépenses. Dans le choix du donneur, on tient compte également du fait que le receveur doit satisfaire certaines règles de cohérence après l'imputation de ces données manquantes. Pour la plupart des sections, l'imputation est effectuée au niveau du ménage mais dans certains cas, par exemple pour le revenu et les dépenses en vêtements, l'imputation est effectuée au niveau de la personne car c'est à ce niveau que l'on collecte l'information pour ces variables.

Notons que depuis 2001, l'imputation des données sur le revenu et les dépenses est effectuée à l'aide du Système canadien de contrôle et d'imputation du recensement (SCANCIR) de Statistique Canada. Le nouveau système utilise une méthodologie quelque peu différente de celle du système utilisé auparavant. Ce système permet une meilleure utilisation des données catégoriques lors de la sélection du donneur. De plus, ce système permet à la fois l'imputation des données continues et des données catégoriques. Des tests ont été effectués avant l'introduction du système et les résultats obtenus étaient semblables à ceux obtenus avec l'ancien système. À compter de 2003, les données catégoriques, qu'on trouve principalement dans les sections sur les caractéristiques et l'équipement du logement, sont imputées à l'aide du SCANCIR. Les données catégoriques étaient imputées auparavant à l'aide d'une technique "hot deck" où un ménage donneur est choisi de façon aléatoire parmi un groupe de ménages répondants possédant des caractéristiques semblables.

Le biais des estimations causé par l'imputation de la non-réponse partielle est difficile à évaluer. Il dépend de la différence entre les répondants et les non-répondants ainsi que de la capacité de la méthode d'imputation à produire des estimations non biaisées. Toutefois, les taux d'imputation donnent une indication de l'importance des non-réponses partielles. Ils sont présentés dans la section suivante.

5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale, provinciale et territoriale

Une première indication de l'ampleur de la non-réponse partielle est la proportion de ménages nécessitant imputation et le nombre de variables imputées par ménage. Le questionnaire peut se diviser en deux grands groupes de variables, celles qui sont recueillies au niveau du ménage et celles collectées auprès des individus telles que le revenu et les dépenses en vêtements. Pour ces dernières catégories, il est important de noter que l'on accepte que le répondant fournisse seulement le revenu total ou les dépenses totales s'il est incapable de fournir la ventilation par source de revenu ou type de dépenses. Le niveau d'imputation pour les composantes du revenu et des dépenses vestimentaires est donc plus élevé mais n'affecte pas le revenu total, les dépenses vestimentaires totales ainsi que le total des dépenses.

Le pourcentage de ménages pour lesquels il a fallu imputer des dépenses de ménage (excluant les dépenses vestimentaires et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent) est présenté dans la prochaine

sous-section. Dans celle qui la suit, on présente le pourcentage de personnes pour lesquelles il a fallu imputer au moins une variable reliée aux dépenses vestimentaires, le pourcentage de personnes pour lesquelles au moins une variable de revenu a été imputée et le pourcentage de personnes pour lesquelles il a fallu imputer au moins une variable de la section sur l'impôts personnels, sécurité et dons en argent. Finalement, dans la dernière sous-section, on présente les résultats pour le pourcentage de ménages pour lesquels il a fallu imputer au moins une des variables catégoriques. À la suite de l'imputation des données par le système, certaines corrections peuvent s'avérer nécessaires, autant sur les variables imputées par le système que sur les autres, afin d'assurer la cohérence des données. Dans les faits, ces modifications représentent un très faible pourcentage. Les résultats sont fournis à l'échelle nationale, provinciale et territoriale. Cela fournit une indication des provinces ou territoires qui sont les plus affectés par l'imputation.

5.1.1 Imputation des dépenses des ménages par province et territoire

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable de dépenses figure dans le tableau 5.1-1. Les ménages utilisables correspondent à tous les ménages demeurant dans des logements admissibles, à l'exception de ceux qui n'ont pu être contactés, qui ont refusé de participer à l'enquête ou qui ont fourni des données incomplètes (voir définitions dans la section 2.1). Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 259) pour un ménage.

Il importe de préciser que les paiements hypothécaires réguliers ainsi que les primes d'assurance hypothécaire sont inclus au titre des dépenses de logement et donc dans les dépenses totales. Depuis 2002, ces deux variables sont ajoutées au calcul des taux d'imputation présentés au tableau 5.1-1. Ceci a pour effet d'augmenter le taux global d'imputation.

Depuis 2004, un changement a été apporté au questionnaire en ce qui concerne les dépenses reliées aux services de communication au foyer (téléphone, téléphone cellulaire et accès à Internet), aux services de télédistribution, aux services de diffusion par satellite ainsi qu'aux systèmes de sécurité. En raison de l'utilisation grandissante des forfaits (services groupés), un ménage peut-être facturé pour des services combinés, de façon qu'il est impossible pour celui-ci de fournir les dépenses pour les services individuels. On accepte alors que le ménage répondant fournisse seulement les dépenses totales de ces services, tout en indiquant quels services sont inclus dans le forfait. L'imputation des dépenses des services individuels s'effectue alors en deux étapes. Dans un premier temps on impute les ménages pour lesquels seulement quelques services sont manquants, suivis des ménages pour lesquels seulement le total des dépenses du forfait est disponible. Pour ces derniers ménages, les dépenses des services (ceux inclus dans le forfait) imputés sont ajustées au prorata afin que leur somme corresponde au total des dépenses du forfait fourni par le ménage répondant. Étant donné que ce changement a eu un impact important sur le taux global d'imputation des dépenses, les taux d'imputation du tableau 5.1-1 sont présentés séparément avec et sans les frais de services de communication au foyer, de location de services de télédistribution, de location de services de diffusion par satellite et de location de systèmes de sécurité. De plus, puisque ce

changement a eu un impact sur le niveau d'imputation des dépenses de ces six services, le taux d'imputation ainsi qu'une mesure d'impact de l'imputation pour chacun de ces services sont présentés au tableau 5.1-2.

Tableau 5.1-1
Ménages qui nécessitent l'imputation de dépenses par province et territoire ainsi qu'à l'échelle nationale

	Ménages qui nécessitent l'imputation de **variables de dépenses**[1] (sauf pour les dépenses en vêtements et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent)							
	Exclut les frais de services de communication au foyer et les frais de location pour télédistribution (câble), pour la diffusion par satellite et pour les systèmes de sécurité				Inclut les frais de services de communication au foyer et les frais de location pour télédistribution (câble), pour la diffusion par satellite et pour les systèmes de sécurité			
	Nombre de variables imputées (parmi 259)				Nombre de variables imputées (parmi 265)			
	1	2	3 ou plus	Total	1	2	3 ou plus	Total
	%							
Canada	14,9	6,4	7,6	29,0	7,5	17,7	33,7	58,9
Terre-Neuve-et-Labrador	15,1	4,1	5,0	24,1	6,6	17,7	30,7	55,0
Île-du-Prince-Édouard	17,3	7,2	4,9	29,3	7,4	17,7	42,7	67,8
Nouvelle-Écosse	15,3	7,3	7,2	29,8	5,9	17,9	43,5	67,3
Nouveau-Brunswick	18,4	6,8	7,1	32,3	9,5	16,7	32,7	59,0
Québec	18,8	6,2	9,7	34,7	9,1	13,9	39,8	62,8
Ontario	12,7	7,8	8,3	28,7	7,7	14,2	34,3	56,2
Manitoba	14,6	7,7	8,2	30,5	8,1	16,3	30,3	54,7
Saskatchewan	15,0	6,6	7,0	28,6	5,6	20,3	39,6	65,6
Alberta	11,7	5,5	8,6	25,8	6,7	21,2	31,1	59,1
Colombie-Britannique	13,5	7,0	7,9	28,4	6,3	25,5	30,0	61,7
Yukon	17,6	3,3	5,1	26,1	12,1	21,7	14,0	47,8
Territoires du Nord-Ouest	10,4	3,0	14,3	27,8	7,8	13,9	17,8	39,6
Nunavut	8,6	3,7	2,1	14,4	8,6	4,3	3,2	16,0

1. Inclut les paiements hypothécaires réguliers et les primes d'assurance hypothécaire.

Puisque le niveau d'imputation des dépenses reliées aux services de communication au foyer, aux services de télédistribution, aux services de diffusion par satellite et aux systèmes de sécurité est plus élevé, il est important de mesurer l'effet de l'imputation sur les estimations de totaux de ces six variables. Cette mesure ainsi que le taux d'imputation permettront de faire un suivi dans le temps de l'ampleur de l'imputation effectuée pour ces variables. La popularité grandissante des forfaits (services groupés) dans la population fait en sorte que le niveau d'imputation devrait augmenter au cours du temps. La mesure d'impact de l'imputation est définie par le total pondéré des données imputées divisé par l'estimation totale (sommes des valeurs pondérées). Elle représente la proportion de la valeur totale de l'estimation provenant de données imputées.

Tableau 5.1-2
Impact de l'imputation à l'échelle nationale des frais de services de communication au foyer, des frais de location pour la télédistribution, pour la diffusion par satellite et pour les systèmes de sécurité

	Impact de l'imputation	Ménages déclarants qui nécessitent l'imputation des frais de services de communication au foyer, des frais de location pour télédistribution (câble), pour la diffusion par satellite ou pour des systèmes de sécurité
	%	
Services téléphoniques	34,7	40,8
Services de téléphones cellulaires, téléavertiseurs et appareils de messagerie textuelle portatifs	11,5	13,2
Location de services de télédistribution (câble)	43,8	44,5
Location de services pour la diffusion par satellite	14,4	11,8
Services d'accès à Internet	52,8	58,5
Systèmes de sécurité	4,5	7,6

5.1.2 Imputation des dépenses des personnes et du revenu par province et territoire

Étant donné qu'une partie des répondants fournissent uniquement le total pour les variables de dépenses en vêtements et les variables de revenu, celles-ci sont imputées, au niveau des individus, en deux étapes. Les individus pour lesquels il faut imputer seulement quelques composantes sont imputés en premier lieu, suivis par ceux pour lesquels seulement les totaux sont disponibles et qui nécessitent l'imputation de toutes les composantes. (Voir la référence [1] pour une description plus détaillée de ce processus).

Le pourcentage d'individus utilisables (personnes membres des ménages utilisables) pour lesquels il a fallu imputer au moins une variable de revenu est indiqué par province et territoire dans le tableau 5.2. On y retrouve le pourcentage de personnes pour lesquelles exactement une variable a été imputée, ceux dont deux variables ou plus (mais pas toutes) ont été imputées et le pourcentage des personnes pour lesquelles seul le revenu total était disponible, nécessitant ainsi l'imputation de toutes les composantes. Le pourcentage total de personnes pour lesquelles une forme d'imputation de revenu a été effectuée est également donné. L'avant dernière colonne du tableau 5.2 contient ce même pourcentage total de personnes pour lesquelles une forme d'imputation a été effectuée, mais pour les dépenses en vêtements. La dernière colonne contient le pourcentage total de personnes pour les lesquelles une forme d'imputation a été effectuée pour les variables de la section du questionnaire portant sur les impôts personnels, sécurité et dons en argent.

Il importe de préciser que seuls les membres du ménage âgés de 15 ans et plus au 31 décembre de l'année de référence doivent répondre aux questions portant sur le revenu personnel et celles de la section sur les impôts personnels, sécurité et dons en argent. Par conséquent, depuis l'année de référence 2003, le pourcentage de personnes qui nécessitent l'imputation de variables de revenu ou pour celles de la section sur l'impôts personnels, sécurité et dons en argent a été calculé en se restreignant aux personnes de 15 et plus et non sur l'ensemble des personnes comme pour les années précédentes. Ceci a pour effet d'augmenter légèrement le taux d'imputation pour ces variables. Tout comme pour les années précédentes, le pourcentage de personnes qui nécessitent l'imputation de dépenses vestimentaires est calculé sur l'ensemble des personnes, puisque tous les membres d'un ménage doivent répondre aux questions portant sur ce type de dépenses.

Tableau 5.2
Personnes qui nécessitent l'imputation du revenu, personnes qui nécessitent l'imputation de dépenses en vêtements et personnes qui nécessitent l'imputation de variables de la section sur les impôts personnels, sécurité et dons en argent par province et territoire ainsi qu'à l'échelle nationale

	Pourcentage de personnes qui nécessitent l'imputation de variables de revenu				Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 15 variables de dépenses vestimentaires	Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 15 variables de la section sur les impôts personnels, sécurité et dons en argent
	1 variable imputée	2 variables imputées ou plus (pas toutes)	Toutes les variables imputées (revenu total connu)	**Total** (toute forme d'imputation)		
	%					
Canada	4,6	2,1	0,6	7,5	10,3	16,0
T.-N.-L.	2,6	1,0	0,7	4,4	12,1	9,6
Î.-P.-É.	4,5	4,1	1,0	10,0	12,8	18,0
N.-É.	4,0	2,9	0,9	7,9	17,2	19,7
N.-B.	3,8	1,3	0,5	5,9	7,7	17,9
Qc	6,2	2,2	0,2	8,6	9,6	16,1
On.	5,2	2,3	0,3	7,8	6,1	17,3
Man.	7,0	2,5	0,3	9,9	6,5	16,4
Sask.	4,6	2,5	0,8	8,0	7,4	15,7
Alb.	4,3	2,0	0,6	7,1	7,0	16,8
C.-B.	5,1	2,5	1,0	8,7	18,4	18,7
Yukon	4,7	0,6	0,4	5,7	9,6	11,2
Territoires du Nord-Ouest	1,7	0,8	0,0	2,7	21,3	8,5
Nunavut	1,8	0,9	0,2	2,9	5,8	5,4

5.1.3 Imputation des variables catégoriques par province et territoire

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable catégorique figure dans le tableau 5.3. Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 42) pour un ménage. Les variables catégoriques qui sont imputées se trouvent aux sections suivantes du questionnaire: Caractéristiques du logements (à l'exception de la variable genre de logement), Équipement associé au logement, Mode d'occupation du logement (à l'exception des variables se rapportant au changement dans le mode d'occupation durant l'année de référence), Tabac et dépenses diverses, pour les variables se rapportant aux achats par le biais de vente directe (questions pour lesquelles la réponse est oui ou non). Il est à noter que les autres variables catégoriques du questionnaire telles que celles se rapportant à la composition du ménage ou encore aux divers branchements du questionnaire sont vérifiées et validées par les spécialistes du sujet de la Division de la statistique du revenu. Par conséquent, ces dernières variables ne sont pas imputées au moyen de la technique du plus proche voisin.

Tableau 5.3
Ménages qui nécessitent l'imputation de variables catégoriques par province et territoire ainsi qu'à l'échelle nationale

	Nombre de variables imputées (parmi 42)			
	1	2	3 ou plus	Total
	%			
Canada	7,2	1,6	1,0	9,8
Terre-Neuve-et-Labrador	4,1	0,1	0,1	4,3
Île-du-Prince-Édouard	10,5	0,8	1,7	13,0
Nouvelle-Écosse	6,0	1,4	0,4	7,9
Nouveau-Brunswick	6,2	1,6	0,0	7,8
Québec	6,8	1,1	0,8	8,7
Ontario	6,9	2,2	2,5	11,6
Manitoba	10,2	3,4	0,9	14,6
Saskatchewan	6,1	2,0	0,8	8,9
Alberta	6,9	1,3	0,6	8,8
Colombie-Britannique	8,4	1,7	1,2	11,3
Yukon	8,8	0,7	1,5	11,0
Territoires du Nord-Ouest	16,1	5,2	2,6	23,9
Nunavut	4,3	0,0	0,5	4,8

Références

[1] Tremblay J. et Arsenault, S. 2001. *La méthodologie de l'Enquête sur les dépenses des ménages.* N° 62F0026MIF2001003 au catalogue. Ottawa. Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.

[2] Rao, J.N.K., Wu, C.F.J. et Yue K. 1992. « Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes », *Techniques d'enquête*, 18(2), p.225-234 (Statistique Canada, N° 12-001 au catalogue).

[3] Statistique Canada, Division de la statistique du revenu. 2009. *Guide de l'utilisateur — Enquête sur les dépenses des ménages.* N° 62F0026M au catalogue – N° 001. Ottawa.

[4] Statistique Canada, Division de la statistique du revenu. 2009. *Les habitudes de dépenses au Canada.* N° 62-202 au catalogue. Ottawa.

[5] Lessard, S. 2005. *Révision de la stratégie de calage de l'Enquête sur les dépenses des ménages.* Ottawa. Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.

Annexe A

Notation algébrique

1. Ajustement pour la non-réponse

Le sous-poids (c-à-d. le poids de sondage ajusté pour la non-réponse) d'un ménage k, noté w_k^{NR}, est

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{taux_g} \quad avec \quad taux_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

où

$s_{g,r}$ est l'ensemble des ménages répondants dans le groupe de non-réponse g,

$s_{g,nr}$ est l'ensemble des ménages non-répondants (refus, unités non contactées, données inutilisables) dans le groupe de non-réponse g, et

π_k^{-1} est le poids de sondage attribué au ménage k.

2. Calcul du taux de glissement

Le taux de glissement pour un groupe de contrôle c, noté $taux_c$, est,

$$taux_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR}\right) - t_c}{t_c}$$

où

$s_{c,r}$ est l'ensemble des ménages répondants dans le groupe de contrôle c,

w_k^{NR} est le sous-poids du ménage k, et

t_c est le total de la donnée auxiliaire pour le groupe de contrôle c.